

A novel representation of rank constraints for non-square real matrices

Ramón A. Delgado, Juan C. Agüero and Graham C. Goodwin
School of Electrical Engineering and Computer Science,
The University of Newcastle, Australia

October 2014

Abstract

We present a novel representation of rank constraints for non-square real matrices. We establish relationships with some existing results, which are particular cases of our representation. One of these particular cases, is a representation of the ℓ_0 pseudo-norm, which is used in sparse representation problems. Finally, we describe how our representation can be included in rank-constrained optimization and in rank-minimization problems.

Notation and basic definitions: $\text{rank}(A)$ denotes the rank of a matrix A . We denote by A^\dagger to the Moore-Penrose pseudoinverse of A . $\lambda_i(A)$ denotes the i -th largest eigenvalue of a symmetric matrix A , $A \succeq 0$ denotes that A is positive semidefinite, and $A \succeq B$ denotes that $A - B \succeq 0$. We represent the transpose of a given matrix A as A^\top . \mathbb{S}^n denotes the set of symmetric matrices of size $n \times n$, and \mathbb{S}_+^n the set of symmetric positive semidefinite matrices, i.e. $\mathbb{S}_+^n := \{A \in \mathbb{S}^n | A \succeq 0\}$. $\|\cdot\|_F$ denotes the Frobenius norm. $\|\cdot\|_0$ denotes the ℓ_0 pseudo-norm that counts the number of nonzero elements of a vector.

1 Main Result

Theorem 1 *Let $G \in \mathbb{R}^{m \times n}$, then the following expressions are equivalent*

- (i) $\text{rank}(G) \leq r$
- (ii) $\exists W_R \in \Phi_{n,r}$, such that $GW_R = 0_{m \times n}$
- (iii) $\exists W_L \in \Phi_{m,r}$, such that $W_L G = 0_{m \times n}$

where

$$\Phi_{n,r} = \{W \in \mathbb{S}^n, 0 \preceq W \preceq I, \text{trace}(W) = n - r\} \quad (1)$$

Proof: Here we provide a sketch of the proof. A more detailed proof will be published somewhere else.

We first prove (i) \implies (ii). Let $\text{rank}(G) \leq r$ then there exists at least $n - r$ linearly independent vectors $u_i \in \mathbb{R}^n$ such that $Gu_i = 0$. Define $U =$

$[u_1, \dots, u_{n-r}] \in \mathbb{R}^{n \times (n-r)}$ having full column rank. Then we can construct a orthogonal projector, $W_R = UU^\dagger$ which satisfies the condition $\text{rank}(W_R) = n-r$ and is such that $GW_R = 0$. Since W_R is an orthogonal projector it also satisfies $W_R \in \mathbb{S}^n$, $0 \preceq W_R \preceq I_n$ and $\text{rank}(W_R) = \text{trace}(W_R) = n-r$, i.e. $W_R \in \Phi_{n,r}$.

The procedure to prove **(i)** \implies **(iii)** is similar to the proof **(i)** \implies **(ii)**.

Next, we prove **(ii)** \implies **(i)**. For all $W_R \in \mathbb{S}^n$ such that $0 \preceq W_R \preceq I$, it is true that

$$\text{trace}(W_R) \leq \text{rank}(W_R) \quad (2)$$

On the other hand, by using *Sylvester's Inequality* (see e.g. [Bernstein, 2009, Proposition 2.5.9]), we have that

$$\text{rank}(G) + \text{rank}(W_R) \leq n + \text{rank}(GW_R) \quad (3)$$

Then, by using (2), we have

$$\text{rank}(G) + \text{trace}(W_R) \leq n + \text{rank}(GW_R) \quad (4)$$

Then by using the fact that $\text{rank}(GW_R) = \text{rank}(0_{m \times n}) = 0$ we obtain

$$\text{rank}(G) \leq n - \text{trace}(W_R) \quad (5)$$

Since $W_R \in \Phi_{n,r}$, we have that $\text{trace}(W_R) = n-r$. Then

$$\text{rank}(G) \leq r \quad (6)$$

This completes the proof that **(ii)** \implies **(i)**. The procedure to prove **(iii)** \implies **(i)** is similar to the proof **(ii)** \implies **(i)**. \square

To the best of the authors' knowledge, Theorem 1 is novel. The closest results in rank-constrained optimization, is described in [Markovsky, 2014, Markovsky, 2012b] where the rank-nullity theorem is used to establish that, for a matrix $G \in \mathbb{R}^{m \times n}$,

$$\text{rank}(G) \leq r \iff \exists \text{ a full row rank matrix } U \in \mathbb{R}^{(m-r) \times m} \text{ such that } UG = 0 \quad (7)$$

However, requiring that U is full row rank is not easy. For example, it may lead to the necessity of including additional non-convex constraints, such as $UU^\top = I_{m-r}$.

Another closely related result is described in [Dattorro, 2005, §4.4]. The latter result make use of the convex set $\Phi_{n,r}$, but the formalism is valid only for positive semidefinite matrices. The above result establishes that for a matrix $G \in \mathbb{S}_+^n$,

$$\text{rank}(G) \leq r \iff \exists W \in \Phi_{n,r} \text{ such that } \text{trace}(WG) = 0. \quad (8)$$

Notice that Theorem 1 can be seen as a generalisation of (8).

There exist other rank-constraint representations which impose conditions on the coefficients of the characteristic polynomial of the matrix, see [d'Aspremont, 2003, Helmerrsson, 2009]. These representations are valid only for positive semidefinite matrices.

Notice that one of the key steps in proving Theorem 1 is the observation that for all $W \in \mathbb{S}^n$ such that $0 \preceq W \preceq I$, it is true that $\text{trace}(W) \leq \text{rank}(W)$. This fact is a consequence of a stronger result that says that in the set of interest, $\{W \in \mathbb{S}^n | 0 \preceq W \preceq I\}$, the trace function is the largest convex function that is less than or equal to the rank function. This latter result is one of the key underlying ingredients in the development of the nuclear norm heuristic [Fazel, 2001].

In the remainder of this section we establish connection between Theorem 1 and other existing results. The following lemma establishes the relationship between Theorem 1 and the rank-constraint representation in (8).

Lemma 1 *Let $G \in \mathbb{S}_+^n$ and $W \in \mathbb{S}_+^n$, then*

$$\text{trace}(WG) = 0 \iff WG = 0 \quad (9)$$

Proof: Since G and W are symmetric and positive semidefinite, then by the *Cholesky decomposition*, see e.g. [Bernstein, 2009, Fact 8.9.37], there exist matrices $P \in \mathbb{R}^{n \times n}$ and $Q \in \mathbb{R}^{n \times n}$ such that

$$G = PP^\top \quad (10)$$

$$W = QQ^\top \quad (11)$$

We then have that

$$\text{trace}(WG) = \text{trace}(QQ^\top PP^\top) \quad (12)$$

$$= \text{trace}(Q^\top PP^\top Q) \quad (13)$$

Next, we recall that for $A \in \mathbb{R}^{m \times n}$ the *Frobenius norm* is defined by $\|A\|_F = \sqrt{\text{trace}(A^\top A)}$, see e.g. [Bernstein, 2009, page 547]. Then, we have

$$\text{trace}(WG) = \text{trace}(Q^\top PP^\top Q) = \|P^\top Q\|_F^2 \quad (14)$$

and from the definition of a norm we have that $\|A\| = 0$ if and only if $A = 0$, see e.g. [Bernstein, 2009, Definition 9.2.1.]. Then we have that

$$\text{trace}(WG) = \|P^\top Q\|_F^2 = 0 \implies WG = 0 \quad (15)$$

This concludes the proof for $\text{trace}(WG) = 0 \implies WG = 0$. The proof for $WG = 0 \implies \text{trace}(WG) = 0$ is straightforward. \square

Another particular case of Theorem 1 is the representation of the ℓ_0 pseudo norm, denoted as $\|\cdot\|_0$, which corresponds to the number of non-zero elements of a vector. The connection is made by considering a diagonal matrix $G \in \mathbb{R}^{n \times n}$ such that its diagonal elements are given by a vector $x \in \mathbb{R}^n$, i.e. $G = \text{diag}\{x\}$ and we have that $\|x\|_0 = \text{rank}(G)$. Then, Theorem 1 can be used to prove the following result.

Corollary 1 *Let $x \in \mathbb{R}^n$, then the following expressions are equivalent*

- (i) $\|x\|_0 \leq r$
- (ii) $\exists w \in \{w \in \mathbb{R}^n | 0 \leq w_i \leq 1, i = 1, \dots, n; \sum_{i=1}^n w_i = n - r\}$, such that $x_i w_i = 0$ for $i = 1, \dots, n$.

Proof: Consider the following definition $G = \text{diag}\{x\} \in \mathbb{R}^{n \times n}$, i.e.

$$G = \begin{bmatrix} x_1 & 0 & 0 & \cdots & 0 \\ 0 & x_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & x_{n-1} & 0 \\ 0 & \cdots & 0 & 0 & x_n \end{bmatrix} \quad (16)$$

Notice that from construction $\text{rank}(G) = \|x\|_0$. From Theorem 1 we have that $\text{rank}(G) \leq r$, if and only if, there exist a $W \in \{W \in \mathbb{S}^n \mid 0 \preceq W \preceq I; \text{trace}(W) = n - r\}$ such that $GW = 0$. Since G is diagonal, then without loss of generality, we can assume that $W = \text{diag}\{w\}$. This can be easily seen by defining $C = GW$ and considering that W is symmetric. Note that, since G is diagonal, $C_{ij} = G_{ii}W_{ij}$ and $C_{ji} = G_{jj}W_{ji}$. If $G_{ii} = G_{jj} = 0$ for $i \neq j$ then $W_{ij} = W_{ji}$ can take any value, including zero, and still satisfy $C_{ij} = C_{ji} = 0$. If $G_{ii} \neq 0$ then $W_{ij} = W_{ji} = 0$ in order to satisfy that $C_{ij} = 0$. Finally, conditions on w are directly derived from conditions on W . \square

We note in passing that this representation of ℓ_0 constraints is related to the results reported in [Feng et al., 2013, Piga and Tóth, 2013, d’Aspremont, 2003], [Dattorro, 2005, §4.5].

2 Applications in Optimization

In this section we apply Theorem 1 so as to include rank constraints into optimization problems. In the last decade there has been increasing interest on including the rank matrix function into optimization problems. This is motivated by the introduction of the development of the nuclear norm heuristic [Fazel, 2001], which provides a convex relaxation for rank-minimization problems. The nuclear norm heuristic have been shown to be particularly useful on high-dimensional optimization problems. However, has been shown in [Markovsky, 2012a] that there is an inherent loss of performance on the nuclear norm heuristic.

Theorem 1 can be applied to rank-constrained optimization problems by simply replacing the rank constraint by one of the equivalent representations, as follows

$$\begin{aligned} \mathcal{P}_{rco} : \quad & \min_{\theta \in \mathbb{R}^p} f(\theta) \\ & \text{s.t. } \theta \in \Omega \\ & \text{rank}(G(\theta)) \leq r \end{aligned} \quad \equiv \quad \begin{aligned} \mathcal{P}_{rcoequiv} : \quad & \min_{\theta \in \mathbb{R}^p} \min_{W \in \mathbb{S}^n} f(\theta) \\ & \text{s.t. } \theta \in \Omega \\ & G(\theta)W = 0_{m \times n} \\ & W \in \Phi_{n,r} \end{aligned}$$

On the other hand, Theorem 1 can also be applied to rank-minimization problems by using the epigraph representation [Grant and Boyd, 2008], as follows

$$\begin{array}{ll}
\mathcal{P}_{rm} : \min_{\theta \in \mathbb{R}^p} r & \mathcal{P}_{rmequiv} : \min_{\theta \in \mathbb{R}^p} \min_{W \in \mathbb{S}^n} n - \text{trace}(W) \\
\text{s.t. } \theta \in \Omega & \equiv \text{s.t. } \theta \in \Omega \\
\text{rank}(G(\theta)) \leq r & G(\theta)W = 0_{m \times n} \\
& 0 \preceq W \preceq I_n
\end{array}$$

These ideas has been applied by the current authors to rank-constrained optimization problems. For example, in [Aguilera et al., 2014] Corollary 1 has been used to impose ℓ_0 constraints into a Model Predictive Control problem. In [Delgado et al., 2014] the problem of Factor Analysis is considered. In the latter work, Theorem 1 has been used to relax the restrictive assumption that the noise sequences should be uncorrelated. The equivalence for the rank-minimization problem \mathcal{P}_{rm} to the problem $\mathcal{P}_{rmequiv}$ can be seen as a generalisation to non-square matrices of the results presented in [d’Aspremont, 2003].

References

- [Aguilera et al., 2014] Aguilera, R. P., Delgado, R. A., Dolz, D., and Agüero, J. C. (2014). Quadratic MPC with ℓ_0 -input constraint. In *19th IFAC World Congress*, Cape Town, South Africa.
- [Bernstein, 2009] Bernstein, D. S. (2009). *Matrix mathematics: theory, facts, and formulas*. Princeton University Press.
- [d’Aspremont, 2003] d’Aspremont, A. (2003). A semidefinite representation for some minimum cardinality problems. In *42nd IEEE Conference on Decision and Control, 2003*, volume 5, pages 4985–4990.
- [Dattorro, 2005] Dattorro, J. (2005). *Convex optimization and Euclidean distance geometry*. Meboo Publishing, USA.
- [Delgado et al., 2014] Delgado, R. A., Agüero, J. C., and Goodwin, G. C. (2014). A rank-constrained optimization approach: Application to factor analysis. In *19th IFAC World Congress*, Cape Town, South Africa.
- [Fazel, 2001] Fazel, M. (2001). A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the American Control Conference*, pages 4734–4739. IEEE.
- [Feng et al., 2013] Feng, M., Mitchell, J. E., Pang, J.-S., Shen, X., and Wächter, A. (2013). Complementarity formulations of ℓ_0 -norm optimization problems. Technical report, Northwestern University.
- [Grant and Boyd, 2008] Grant, M. C. and Boyd, S. P. (2008). Graph implementations for nonsmooth convex programs. In *Recent advances in learning and control*, pages 95–110. Springer.
- [Helmerrsson, 2009] Helmersson, A. (2009). On polynomial coefficients and rank constraints. Technical report, Linköping University.

- [Markovsky, 2012a] Markovsky, I. (2012a). How effective is the nuclear norm heuristic in solving data approximation problems? In *Proc. of the 16th IFAC Symposium on System Identification*, pages 316–321, Brussels.
- [Markovsky, 2012b] Markovsky, I. (2012b). *Low Rank Approximation: Algorithms, Implementation, Applications*. Communications and Control Engineering. Springer.
- [Markovsky, 2014] Markovsky, I. (2014). Recent progress on variable projection methods for structured low-rank approximation. *Signal Processing*, 96PB:406–419.
- [Piga and Tóth, 2013] Piga, D. and Tóth, R. (2013). An SDP approach for ℓ_0 -minimization: Application to ARX model segmentation. *Automatica*, 49(12):3646–3653.